MIPRO 2014, 26-30 May 2014, Opatija, Croatia

Movie Posters Classification into Genres Based on Low-level Features

Marina Ivasic-Kos, Miran Pobar, Luka Mikec
Department of Informatics
University of Rijeka
Rijeka, Croatia
{ marinai, mpobar}@uniri.hr

***Abstract*— A person can quickly grasp the genre (drama, comedy, cartoons, etc.) from a movie poster, regardless of visual clutter and the level of details. Bearing this in mind, it can be assumed that simple properties of a movie poster should play a significant role in automated detection of movie genres. Therefore, low-level features based on colors and edges are extracted from poster images and used for poster classification into genres.**

In this paper, poster classification is modeled as a multi-label classification task, where a single movie may belong to more than one class (genre). To simplify and solve the multi-label problem, two methods for multi-label data transformation are described and evaluated given the classification results obtained by distance ranking, Naïve Bayes and RAKEL.

Experiments are conducted on a set of 1500 posters with 6 movie genres. Results provide insights into the properties of the discussed algorithms and features.

Keywords—multi-label classification, data transformation method, movie poster

I. INTRODUCTION

With just a glance at a movie poster, a person can comprehend a variety of perceptual and semantic information. Capturing essential information about the movie at once, regardless of the visual complexity, can be experienced while driving around town and spotting a billboard or walking around the shopping center and cursorily glancing at an advertisement for a film. A person can grasp the genre (drama, comedy, cartoons, etc.) from a poster independently of the short observation time, clutter and the variety of details.

The same phenomenon has already been observed in complex real-world scenes, where persons can promptly recognize the basic-level category of the scene (e.g., a street) [1], its spatial layout (e.g., a street with tall vertical blocks on both sides) [2], as well as other global structural information (e.g., a large volume in perspective).

Assuming that one of the goals of the poster is to convey the information about the film (genre, etc.) to potential moviegoers without them paying a lot of attention, the important information for determining the genre might be contained in global low-level features such as dominant color of the whole poster, texture or color histogram. We thus wanted to develop a method to automatically determine the genres of movies using mostly visual low-level features of their posters. The task is made more complex by the fact that most movies belong to more than one genre, for example „Titanic“ belongs to Action, Disaster, Drama and Romance genres, and „Back to the Future“ belongs to Adventure, Comedy, Sci-Fi and Family genres, according to TMDB [3]

Film genre classification from additional supporting promotional material (trailers) has recently received some attention. In paper [4] low-level visual features extracted from movie trailers have been used to classify 100 movies into 4 genres (drama, action, comedy, horror). In [5] GIST, CENTRIST and W-CENTRIST scene features were applied to a collection of temporally-ordered static key frames to yield a feature representation. These features are used as visual vocabulary for genre classification and tested on 1239 movie trailers.

In [6] the same visual features were used as in [4]. Movies were classified into three genres (action, drama, and thriller) which were selected because of their frequency among movies that were played in Taiwan from 2004 to 2006. Some additional genres were grouped together and presented as those three (e.g. drama included comedy and romance while thriller included horror). All these approaches [4-6] simplify the problem considering only single genre per movie and deal with only single label classification problem.

In our approach, multi-label poster classification is proposed. In Section II two kinds of data transformation methods are explained. A set of low-level features, classification methods used for poster classification into movie genres and evaluation of described data transformation methods are presented in Section III. The paper ends with description of experimental setup, experimental results with discussion and directions for future work.

II. DATA TRANSFORMATION METHODS

The goal of classification is to assign class labels from the set of class labels C to examples from the set of samples E. In many problems, an example belongs to a

single class (single-label classification). In binary classification, an example e is classified in one of two possible classes from C, $C = \{C_1, C_2\}$ as in spam classification where a mail message is declared as either spam or not-spam. In multi-class classification an example is classified into one of more classes {C1, C2 … Ck}, for example a cloud can be classified as nimbus, cumulus, cirrus, etc. However, there are many problems where an example e_j should be classified into more than one class, as in our experiment. This is the case of multi-label classification which can be formally expressed as:

$$\exists\, e_j \in E : \varphi(e_j) = \{C_l, C_m\} \cup Z,\ Z \subseteq \cup_{i=1}^{k} C_i,\ l, m \in 1..k, l \neq m \text{ and } \varphi: E \to C.$$

Thus, in the set of examples E, an example e_j exists that the classifier φ maps into at least two classes C_l and C_m.

If a hierarchical structure exists among some of the classes in C so that one class can be inferred from another, then it is the case of hierarchical multi-label classification. For instance in scene classification, if a scene can be classified as beach, outdoor scene and natural scene, the parent class – outdoor scene will be inferred from the child class – beach.

Where the parent-child relationship does not exist between classes, the problem is known as flat multi-label classification. In case of e.g. classifying songs into moods and genres, one class label cannot be inferred from the other, because there is no parent-child relationship between them. For example, the song Paperback Writer by The Beatles can be classified into mood carefree and genre pop while the song Sabotage from Beastie Boys can be classified into mood brash and genre hip-hop according to [7]. The same problem exists in many different applications like in text classification where an article could belong to crime and sports classes, in medical diagnosis, information retrieval etc.

The multi-label classification problem cannot be directly solved by most classification methods which are designed for single-label classification. One approach to tackle a multi-label classification problem is to transform the data so it fits a series of single-label classification [8]. The other approach contains methods that extend specific learning algorithms in order to handle multi-label data directly.

In this paper, a multi-label classification problem is transformed into single-label classification, in two ways.

In the first, each of the examples that had two or more class labels was transformed into a set of ordered pairs. The first element of each pair is the example and the second one is the class label. Thus, if an example $e_j \in E$ can be classified into $Y_j = \{C_l, C_m, \dots, C_r\}$, $Y_j \subseteq C$ then we replace that example with $|Y_j|$ ordered pairs $(e_j, C_l), (e_j, C_m) \dots (e_j, C_r)$. For instance, the movie „Top Gun“ belongs to Action, Romance and War genres, and would be transformed into the set of ordered pairs containing individual genres: {(Top Gun, Action), (Top Gun, Romance), (Top Gun, War)}.

Now, model that maps first element of the ordered pair (example) to the second element of the pair (class label) can be defined. If the classifier assigns a confidence value to each class label, classes whose confidence scores are greater than a set threshold will be assigned to the example. A related approach is ranking, where classes are ranked according to their confidence scores, and top n are mapped to the example. In the following, this data transformation method is referred to as M1.

In the second case each set of class labels $Y_j = \{C_l, C_m, \dots, C_r\}$ is transformed into a new combined class $C_{l,m,\dots,r}$ that is assigned to the example e_j. In that way we create the new set $C' \supseteq C$, by adding the new combined classes into the set C'. Using this problem transformation method "The Top Gun" example would be transformed into the ordered pair containing one combined genre (Top Gun, Action&Romance&War). This data transformation method is hereinafter referred to as M2.

Both problem transformation methods are algorithm independent, so any classification method appropriate for the single-label classification can be used to map the examples to the new class.

III. EXPERIMENTS

In this section we present our experiments conducted on posters and metadata obtained from the TMDB. Our goal is to automatically determine the genres of movies using visual features of their posters.

A. Data and Preprocessing Step

We have used 1500 posters of movies dated from 1990 onwards, selected so that 6 chosen genres (Drama, Action, Animated, Comedy, War and Horror) had 250 examples each. Since each movie can have multiple genres, additional genre labels (e.g. Mystery, Crime, Family, and Romance) were present in the data. These genres were transformed into the chosen 6 and only two labels per movie were kept. The following rules were used to transform the original genre labels that were present in the data into genre labels that were used for classification (Table I):

TABLE I. TRANSFORMATION RULES

Original genre labels	Genre class label
Action, Disaster, Fantasy, Western	Action
Animation	Animation
Comedy	Comedy
Crime, Drama, Romance, Thriller	Drama
Horror	Horror
War, History	War

We applied this preprocessing step to prevent data scarcity, caused by too many genre labels in C' with too few examples. After the mapping, the resulting number of movies of each genre differs from the initial 250.

From 1500 collected posters, 1000 were used for training and 500 for testing.

After the preprocessing step, the genre labels in method M1 correspond to individual genres, and are from the set

$$C = \text{Drama, Action, Animated, Comedy, War, Horror}\}.$$

In the method M2 the genre labels except labels of individual genres include combined genre labels, and are from the set:

$$C' = C \cup \{\text{Drama\&Action, Drama\&Comedy,..., War\&Horror}\}.$$

B. *Features*

Motivated by the way people glance at billboards and still capture enough information about the film we wanted to see if only simple low-level features are enough for automatic poster classification into movie genres.

Intuitively it can be assumed that the variance of color has a strong correlation with respect to genres, so we used more low-level color features and experimentally tested their discriminate ability in terms of genre classification. Before the low-level features were extracted, each poster was proportionally scaled to 200 pixel width and converted to HSV color space. The relationship between the hues with maximal saturation in HSV color space with their corresponding RGB coordinates is illustrated in Fig. 1.

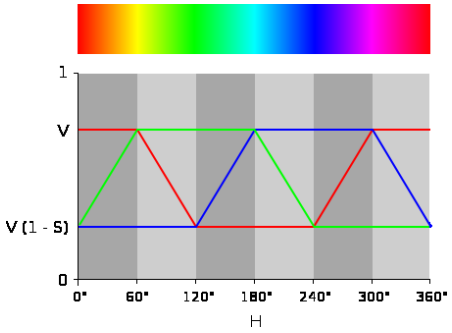


Fig. 1. The relationship between the hue of colors with maximal saturation in HSV color space with their corresponding RGB coordinates [9]

First, the image color histogram was calculated on hue (H) and saturation (S) channels of the whole image. We used 50 bins for the H channel and 60 bins for the S channel. The value (V) channel was not used. The idea was to keep the number of features low enough to emphasize the information that is relevant for classification and adequately train the classifier, especially in approach M2, where the number of classes is increased.

To test if color layout has an important role in classification, we computed 6 local H-S histograms on images divided into 3x2 grids, and one H-S histogram computed on the central part of the image. The central part was of the same proportions as the whole image, but 1/3 of the diagonal size. The arrangement of image grid from which the local color histograms were computed is given in Fig. 2.

Fig. 2. The arrangement of image grid from which the local color histograms were computed

Alternatively, we defined a set of features F which includes dominant colors, features based on edges and the number of faces. Dominant colors correspond to the histogram bins with the highest values and are defined according to [10]. We have used different numbers of dominant colors (3, 6, 8, 12, 16, 24 and 36) and have experimentally determined that 12 dominant colors yield the best classification results in our task.

The edge based feature was computed by first blurring the value (V) channel of each image with the 3 by 3 Gaussian filter, convolving the result with an edge detection mask and counting the number of pixels exceeding the set threshold. Experimentally we selected the value (V) channel of HSV color model and set the threshold to 0.5. We used the Sobel edge detector.

The number of faces in the poster was detected by the OpenCV [11] implementation of Viola-Jones face detector on images blurred with 3 by 3 Gaussian filter. Given that posters are images that are artistically styled, face detection for frontal faces was successful in 66%, while the profiles had significantly worse results, both less than expected for unprocessed images.

C. *Classification Methods*

We used the distance ranking (DR), k-NN and Naïve Bayes classifiers on data adapted with both data transformation methods, and compared the results with the random k-label sets (RAKEL) [12] method which has been shown to be effective in multi-label classification tasks.

For both data transformation methods M1 and M2, we used the same procedure to define representative feature values for the genre labels. For each genre label, we defined a number of representative vectors for each feature except for the number of faces for which the average number was used. The representative values were determined using k-means clustering algorithm. Experimentally, we selected up to 3 representative H-S histograms, 12 dominant colors and 3 centroids of edge based feature values.

For genres that have few movies for one centroids value, less than 3 representative H-S histograms are created. Otherwise genres would be represented with histograms based on too few examples to be relevant. For example, the Horror genre has two representative H-S histograms indicated in the Fig. 3 and Fig. 4, while the Action genre has three representative H-S histograms.

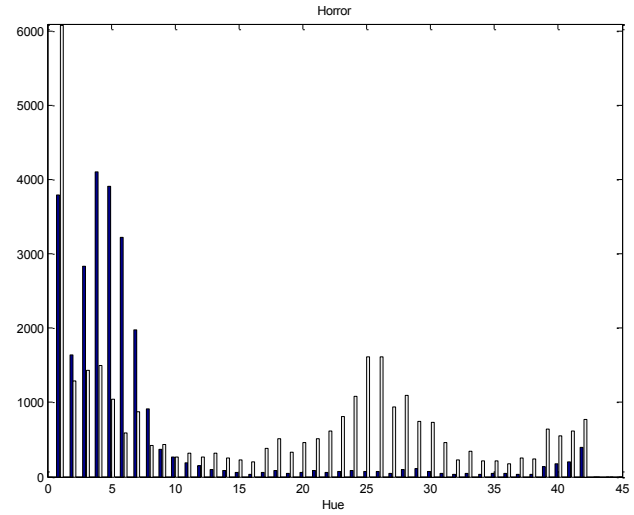


Fig. 3. Representative hue histograms for the Horror genre.

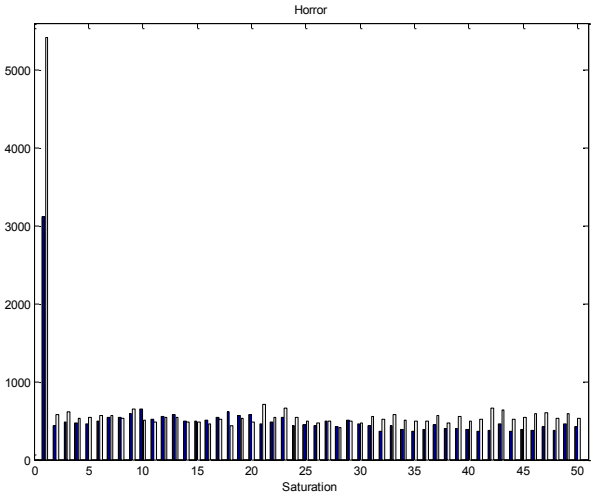


Fig. 4. Representative saturation histograms for the Horror genre.

To determine distance between H-S histograms computed on the poster images, the Chi-Square distance was used (1):

$$d(H_1, H_2) = \sum_I \frac{\left(H_1(I) - H_2(I)\right)^2}{H_1(I)} \qquad (1)$$

where H1, H2 are normalized H-S histograms, I is a histogram bin and d(H1, H2) distance used to express how well both histograms match. For the features from the set F we used the Euclidean distance.

Figures 5 and 6 show the hue and the saturation histograms of Action and Comedy genres. It can be seen for example, that Action posters typically have many more black/gray pixels than Comedy posters, as presented on the saturation histogram. The hue histogram shows the different distribution of colors, e.g. Action genre has more pure red (Hue histogram bin 0) values, while Comedy genre has more orange values (Hue histogram bin 4). Note that hue values range from 0-50 and saturation ranges from 0 to 60.

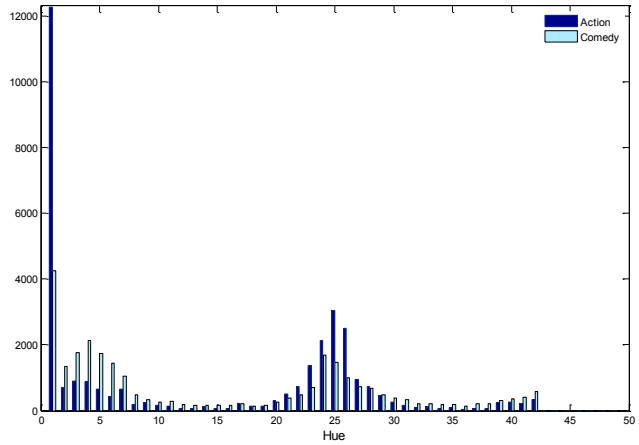


Fig. 5. Action and Comedy hue histograms.

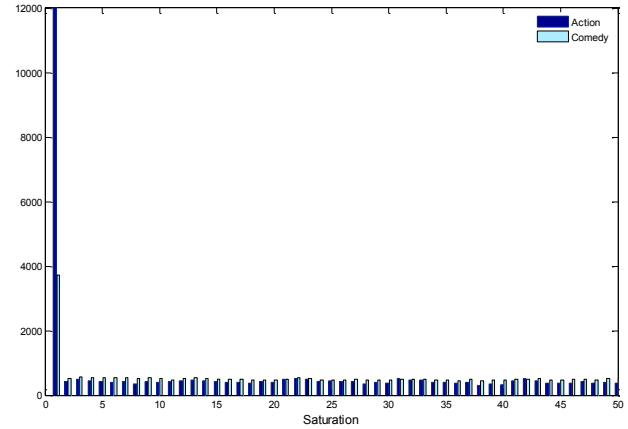


Fig. 6. Action and Comedy saturation histograms.

Moreover, the correlation among dominant colors and movie genres has been experimentally determined. For instance, Animation and Comedy tend to have a large variety of colors (hues), whereas War and Horror films often adopt few dominant colors (hues).

A new unknown movie poster is classified into genres using distance ranking (DR), Naïve Bayes (NB) and RAKEL classification methods.

In the case of DR, the distances of genre centroids and of the feature values extracted from a movie poster are calculated using (1) when histograms are used and Euclidean distance for the features from the set F. The genres are ranked according to the distance of their centroids to the poster features and the top two are selected as classification results.

When using Naïve Bayes, each feature vector x_{e_j} of poster e_j is classified into one of the genre labels according to the maximum posterior probability c_{MAP} (2):

$$c_{MAP} = \text{argmax}_{C_i \in C} P\left(\boldsymbol{x}_{e_j} \middle| C_i\right) P(C_i) \qquad (2)$$

The probability $P(C_i)$ is estimated from relative frequency of label C_i in the training data. To determine the second genre we pick the class with the second highest probability score $P(x_{e_j}|C_i)$.

RAKEL was run using the nearest neighbor (1-NN) classifier based on the Bhattacharyya distance (3),

$$d_B(p,p') = \sum_{i=1}^{N} \sqrt{p(i)p'(i)}, \qquad (3)$$

where p and p' are histograms and N is the number of bins.

The RAKEL subset size was 3 and the number of models was 12.

IV. EXPERIMENTAL RESULTS

All experiments were run using 5 runs of 5-fold cross-validation. The following results were obtained as an average of those runs.

The Table II shows the classification results for the test set of 500 posters using the data transformation approach in which classes correspond to single genres (M1). The presented results are obtained using distance ranking and Naïve Bayes classifier. The results of distance ranking are obtained based on H-S histograms (DR1), local H-S histograms (DR2) or features from the set F (DR3) explained in the Section II. In the Tables II and III, rows marked as "2/2" show the percentage of movies for which both movie genres were correctly recognized and rows marked "1/2" show the percentage of movies for which at least one of the two genres was correct. The best results are obtained using distance ranking based on features from the set F in cases 2/2 and 1/2.Similar results are obtained by Naïve Bayes. Comparing the data, the contribution of color seems to be most important for classification.

TABLE II. CLASSIFICATION RESULTS FOR DATA TRANSFORMATION METHOD M1.

No. correct labels/top ranked labels	DR 1	DR2	DR3	NB
2/2	8.97%	10.10%	**12.1%**	11.2%
1/2	59.72%	54.85%	**66.58%**	62.21%

The results of classification on the same set using the data transformation method with combined genre labels (M2) is presented in the Table III in the similar fashion. In this case, RAKEL method was additionally used.

TABLE III. CLASSIFICATION RESULTS FOR DATA TRANSFORMATION METHOD M2.

No. correct labels/top ranked labels	DR 1	DR 2	DR 3	NB	RAKEL
2/2	**13.61%**	8.90%	10.51%	8.93%	13.24%
1/2	57.69%	58.44%	62.08%	61.24%	**62.54%**

The best results are obtained using distance ranking based on histogram, if all labels need to be correctly classified (2/2). If one correct label should be obtained among the two top ranked (1/2), then RAKEL is the best solution.

Distance ranking based on global H-S histograms used with data transformation method M2 performs best when both genre labels need to be correctly classified. If we allow one incorrect label among the top two ranked, the distance ranking and data transformation method M1 used with the features from the set F (DR3), performed best on test data.

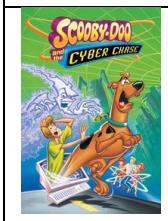
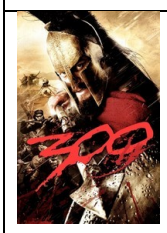
Perhaps surprisingly, results obtained by local H-S histograms are worse in the case of the M2 approach.

The large differences between classification results for cases 2/2 and 1/2 are expected because of the difference in the baseline, where expected random score for 2/2 is 6.7% and for 1/2 is 33%. In both cases, automatic classification performed about twice as good.

The results suggest that both data transformation methods perform similarly regardless of the tested classification method. Consequently, we expect that new features should be extracted from posters to improve the classification.

A few examples of correctly and incorrectly classified posters obtained using DR1 and M2 transformation method are presented in Table IV. Presented movie posters give some insight into the complexity of the problem.

TABLE IV. EXAMPLES OF CORRECTLY AND INCORRECTLY CLASSIFIED POSTERS

Movie poster	No. correct labels/ top ranked labels	Correct labels	Top 3 predicted labels
	2/2	Animation, Comedy	Animation, Comedy, Action
	1/2	Action, War	Action, Horror, Comedy
	0/2	Animation, Comedy	War, Horror, Action

V. Conclusion And Future Work

In this paper, automated detection of movie genres from posters was modeled as a multi-label classification task, where a single movie may belong to more than one genre. The multi-label classification problem was transformed into single-label classification by two kinds of data transformation methods. The first one transforms examples with multiple labels into multiple ordered pairs that are suitable for single label classification. The second extends the set of genre labels by forming combined labels. Both data transformation methods are algorithm independent and were tested with Naive Bayes and distance ranking classifiers and have performed comparably on the test set of 500 movie posters. A well-known multi-label classification algorithm RAKEL was also used with similar results.

The features used in the classification were low-level features based on color and edge combined with the number of detected faces on posters.

The best results were about 14% for two out of two correctly detected labels and 67% for at least one of two correctly detected labels. Since the results were similar using all the tested classification algorithms, we expect that the improvement in classification should primarily come from finding better image features that capture the relevant information about genres in posters. In the future work, we plan to test the dense SURF [13], GIST [14] and other visual low-level features used for scene representation [10] as well as features for text recognition. We also plan to test the classification on a much larger dataset with larger set of genres.

A subjective test will be conducted to determine human ability to detect genres from poster images, and the results will be used for comparison with automatic detection.

References

[1] [1] M. C. Potter, "Short-term conceptual memory for pictures." Journal of Experimental Psychology: Human Learning and Memory, vol. 2, no. 5, p. 509, 1976.

[2] [2] P. G. Schyns and A. Oliva, "From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition," Psychological Science, vol. 5, no. 4, pp. 195–200, 1994.

[3] (2014, 3) The movie database. [Online]. Available: http://www.themoviedb.org/

[4] Z. Rasheed, Y. Sheikh, and M. Shah, "On the use of computable features for film classification," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 15, no. 1, pp. 52–64, 2005.

[5] H. Zhou, T. Hermans, A. V. Karandikar, and J. M. Rehg, "Movie genre classification via scene categorization," in Proceedings of the international conference on Multimedia. ACM, 2010, pp. 747–750.

[6] H.-Y. Huang,W.-S. Shih, and W.-H. Hsu, "A film classifier based on low-level visual features," in Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on. IEEE, 2007, pp. 465–468.

[7] (2014, 3) Allmusic. [Online]. Available: http://www.allmusic.com/

[8] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," Pattern recognition, vol. 37, no. 9, pp. 1757–1771, 2004.

[9] (2014, 3) Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Hue

[10] M. Ivašić-Kos, M. Pavlić, P. Poščić, "Descriptors' Specification for Image Representation", Proceedings of International Conference on Image and Video Processing and Computer Vision. Orlando, Florida.

[11] (2014, 3) OpenCV. [Online]. Available: http://opencv.org

[12] G. Tsoumakas and I. Vlahavas, "Random k-label sets: An ensemble method for multi-label classification," in Machine Learning: ECML 2007. Springer, 2007, pp. 406–417.

[13] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in Computer Vision–ECCV 2006. Springer, 2006, pp. 404–417.

[14] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International journal of computer vision, vol. 42, no. 3, pp. 145–175, 2001.